Hypothesis Testing in Predictive Regressions

Yakov Amihud[1] Clifford M. Hurvich[2] Yi Wang[3]

November 12, 2004

[1]Ira Leon Rennert Professor of Finance, Stern School of Business, New York University, New York NY 10012

[2]Professor of Statistics and Research Professor of Information, Operations and Management Science, Stern School of Business, New York University, New York NY 10012

[3]Ph.D. Student, Operations and Management Science, Stern School of Business, New York University, New York NY 10012

The authors thank Jon Lewellen for helpful comments and suggestions.

Abstract

We propose a new hypothesis testing method for multi-predictor regressions with finite samples, where the dependent variable is regressed on lagged variables that are autoregressive. It is based on the augmented regression method (*ARM*; Amihud and Hurvich (2004)), which produces reduced-bias coefficients and is easy to implement. The method's usefulness is demonstrated by simulations and by an empirical example, where stock returns are predicted by dividend yield and by bond yield spread. For single-predictor regressions, we show that the *ARM* outperforms bootstrapping and that the *ARM* performs better than Lewellen's (2003) method in many situations.

Keywords: Augmented Regression Method (*ARM*); Bootstrapping; Hypothesis Testing.

I Introduction

In a class of predictive regressions analyzed by Stambaugh (1999), a variable is regressed on the lagged value of a predictor variable, which is autoregressive with errors that are correlated with the errors of the regression model. Stambaugh (1999) shows that in finite samples, the estimated predictive slope coefficient is biased, leading to the incorrect conclusions that the lagged variable has predictive power while in fact it does not. Stambaugh (1999) derives the bias expression which is later used in empirical studies to obtain a reduced-bias point estimate of the predictive coefficient.

For hypothesis testing in predictive regressions, three methods have been employed: (1) bootstrapping, used by Kothari and Shanken (1997); (2) testing under an assumption that the autoregressive coefficient is almost 1.0, setting it to be, for example, 0.9999 (Lewellen (2003)); and (3) a method to derive the standard error of the estimated predictive coefficient following an augmented regression method (ARM), in which the predictive regression is estimated using the predictor variable and its bias-adjusted autoregressive residual (Amihud and Hurvich (2004)). Hitherto, the performance of these testing methods has not been fully evaluated.

The first task of this paper is to compare the performance of these three methods of hypothesis testing in terms of the size and power of the test. We find through simulations that the ARM performs quite well. The actual size under the ARM is considerably closer to the nominal size (the ordinary 1%, 5% and 10%) than the size under the other two methods. Method (2) produces accurate size only if the true autoregressive coefficient has the assumed value, but the size becomes inaccurate if there is even a very small difference between the two in two-sided tests. The ARM also has better power than the bootstrapping method.

For multi-predictor regressions, there hitherto exists no feasible method of hypothesis testing. Amihud and Hurvich (2004) propose a reduced-bias estimator of the predictive coefficients, based on ARM. In this paper, we propose a new method for hypothesis testing in multi-predictor regressions, which can be viewed as an extension of the hypothesis testing method proposed for the single-predictor regressions, whose performance is evaluated in the first part of the paper.

The second task of this paper is to propose a convenient new method of hypothesis testing in multi-predictor regressions, and to examine its performance. We first present the theory underlying the proposed method, and then we perform simulations to compare the size under this method with the nominal size. While the differences between the actual and nominal sizes are not as small as in the single-predictor model, they are still reasonably small. Still, our ARM is the only one available for estimation and hypothesis testing in general *multi*-predictor regressions.

Alternatively, there is a local-to-unity asymptotic approach to the predictive regression problem, which allows for a more general error structure, including short-run dynamics (the predictor model is AR(p) rather than AR(1)) and non-normality. Campbell and Yogo (2004) develop a feasible Q-statistic under the local-to-unity framework. A simple pretest is also suggested to determine whether the conventional t-test gives correct inference. Maynard and Shimotsu (2004) suggest a new covariance-based test of orthogonality in the case where the predictors have roots close to or equal to unity. The asymptotic properties are derived and simulations are performed against various reasonable alternatives. Kernel estimation is used for estimation. Our paper, instead, focuses on *finite*-sample properties, under the normality assumption, for estimation methods that do not require the use of kernel methods.

Our paper proceeds as follows. In Section II, we discuss hypothesis testing for single-predictor regressions using three methods: bootstrapping, Lewellen's (2003) method and the ARM. Their performance is compared by simulations. An empirical example is studied using the three methods, which yield different conclusions. Section III describes hypothesis testing in the multi-predictor case, based on a newly proposed estimator of the covariance matrix of the estimated slope coefficients. Both individual and joint tests are suggested and investigated. Again simulations and empirical data analysis are performed to demonstrate the methodology. We present our conclusions in Section IV. Proofs of the theoretical results are presented in the Appendix, Section V.

II Single-Variable Predictive Regression

Consider a single-variable predictive regression model (following Stambaugh (1999)), where a scalar time series $\{y_t\}_{t=1}^{n}$ is to be predicted from a scalar first-order autoregressive ($AR(1)$) time series $\{x_t\}_{t=0}^{n-1}$. The model is,

$$y_t = \alpha + \beta x_{t-1} + u_t \quad , \tag{1}$$

$$x_t = \theta + \rho x_{t-1} + v_t \quad , \tag{2}$$

where the errors (u_t, v_t) are each serially independent and identically distributed as bivariate normal:

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{iid}{\sim} N(0, \Sigma) \quad , \quad \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \quad .$$

The autoregressive coefficient ρ of $\{x_t\}$ satisfies the constraint $|\rho| < 1$, to ensure stationary of $\{x_t\}$. Stambaugh (1999) shows that if $\sigma_{uv} \neq 0$, the ordinary least squares (OLS) estimator $\hat{\beta}$ based on a finite sample is biased: $E(\hat{\beta} - \beta) = \phi E(\hat{\rho} - \rho)$, where $\phi = \sigma_{uv}/\sigma_v^2$.

In applications that followed, researchers estimate β as $\hat{\beta}^s = \hat{\beta} + (\hat{\sigma}_{\hat{u}\hat{v}}/\hat{\sigma}_{\hat{v}}^2)(1 + 3\hat{\rho})/n$, where hat indicates an OLS estimated parameter or variable, and $E(\hat{\rho} - \rho)$ is estimated by $(1 + 3\hat{\rho})/n$ following Kendall (1954).

Hypothesis testing for this model can be performed by several existing methods. The methods we discuss here are the bootstrapping method applied by Kothari and Shanken (1997), Lewellen's (2003) method and the Amihud-Hurvich (2004) method. In what follows, we describe each method and then compare their performances.

A Hypothesis Testing in Single Variable Predictive Regression

A.1 Bootstrapping (BS)

The estimation of the standard error of the predictive coefficient by bootstrapping has been used by Nelson and Kim (1993), Kothari and Shanken (1997) and Baker and Stein (2003). We consider both nonparametric and parametric bootstrapping.

For the nonparametric bootstrapping procedure (BS^N), we follow the version in Kothari and Shanken (1997), which is briefly summarized as follows:

1) Given the time series $\{y_t, x_t\}$, model (1) and (2) are estimated by OLS to get estimated parameters, $\hat{\alpha}, \hat{\beta}, \hat{\theta}$ and $\hat{\rho}$, as well as the residuals $\{\hat{u}_t\}, \{\hat{v}_t\}$.

2) The estimated parameters are corrected for bias using,

$$\begin{aligned} \hat{\rho}_A &= \frac{n\hat{\rho}+1}{n-3} \\ \hat{\beta}_A &= \hat{\beta} + (\frac{\hat{\sigma}_{\hat{u}\hat{v}}}{\hat{\sigma}^2_{\hat{v}}})(\frac{1+3\hat{\rho}_A}{n}) \end{aligned}$$

where n is the sample size.

3) Given $\hat{\beta}_A, \hat{\rho}_A, \hat{\alpha}, \hat{\theta}$ and $\{y_t, x_t\}$, calculate the residuals $\{\hat{u}_{A,t}\}, \{\hat{v}_{A,t}\}$ using equations (1) and (2).

4) The testing of the null hypothesis $H_0 : \beta = \beta_0$ is performed by calculating the empirical p-value (tail area) for the observed $\hat{\beta}_A$, based on the simulated null distribution. Thus we need to simulate a number of replications, and for each one construct a corresponding estimated value $\hat{\beta}^*_A$, in order to obtain the empirical bootstrap distribution for $\hat{\beta}^*_A$.

Each bootstrap replication is constructed by fixing the first observation x_0 (from the data) and constructing $\{y^*_t, x^*_t\}$ iteratively using simulated data generated by the parameter β_0 (the hypothesized value), the estimates $\hat{\rho}_A, \hat{\alpha}$ and $\hat{\theta}$ and equations (1) and (2). The simulated residuals (u^*_t, v^*_t) are selected randomly with replacement from all possible pairs of $(\hat{u}_{A,t}, \hat{v}_{A,t})$ obtained in Step (3). The bootstrap replication is denoted by $\{y_t^*, x_t^*\}_{t=0}^n$. [1]

5) Using the bootstrap replication $\{y_t^*, x_t^*\}$, repeat Steps (1) and (2), obtaining the corrected slope estimate $\hat{\beta}_A^*$. Repeat Step (4) M times (we use M=2500, as in Kothari and Shanken (1997)), fixing β_0 and the starting observation x_0. We obtain M values of $\hat{\beta}_A^*$ for the given data set, from which we generate the empirical bootstrap distribution of $\hat{\beta}_A^*$.

6) Rank these M values of $\hat{\beta}_A^*$ and decide whether to reject the null hypothesis by comparing $\hat{\beta}_A$ to the bootstrap distribution of $\hat{\beta}_A^*$. For example, in a two-sided nominal size α test of $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$, reject the null hypothesis if $\hat{\beta}_A$ is smaller than the $(\alpha/2)$ quantile or larger than the $1 - (\alpha/2)$ quantile of $\hat{\beta}_A^*$. Similar calculations are done for the one-sided test of $H_0 : \beta = \beta_0$ against $H_a : \beta > \beta_0$.

7) We examine the performance of these bootstrap tests over 1500 simulated data sets $\{y_t, x_t\}$.

The parametric bootstrapping procedure[2] is similar to the nonparametric one, except that in Step (4), the bootstrap errors (u_t^*, v_t^*) are simulated from a bivariate normal distribution with covariance matrix estimated from $\{\hat{u}_{A,t}, \hat{v}_{A,t}\}$. As for the first observation x_0, two alternative methods (referred as *parametric-fixed* (BS^{Pf}) and *parametric-random* (BS^{Pr}) bootstrapping hereafter) can be used: either fix x_0 from the data or draw it randomly from its estimated distribution $N\left[\frac{\hat{\theta}}{1-\hat{\rho}_A}, \frac{\widehat{Var}(\hat{v}_A)}{1-(\hat{\rho}_A)^2}\right]$. Here, a technical problem arises: if ρ is close to 1, the Kendall (1954) correction could lead to $|\hat{\rho}_A| > 1$, and the

[1] The two intercepts $\hat{\alpha}$ and $\hat{\theta}$ are not corrected because they have no effect on the bias of $\hat{\beta}$.
[2] This procedure is used by Polk, Thompson and Vuolteenaho (2004).

corresponding variance would be negative. In this case, x_0 is set to be the estimated mean value: $\frac{\hat{\theta}}{1-\hat{\rho}_A}$.

A.2 Lewellen's Method (L)

Lewellen (2003) proposes a hypothesis testing method based on the empirical observation that the autoregressive coefficient ρ is very close to unity in some financial time series. Assuming $\rho \approx 1$ (e.g., 0.9999), an upper bound for the bias in $\hat{\beta}$ is estimated and the one-sided hypothesis test based on the corresponding bias-corrected version of $\hat{\beta}$ may be regarded as conservative. Lewellen demonstrates the improvement in the power of the test under the assumption $\rho \approx 1$. The following is a brief description of the method.

1) Perform OLS regression using equations (1) and (2) to get estimated parameters $\hat{\beta}$ and $\hat{\rho}$ and residual series $\{\hat{u}_t\}$ and $\{\hat{v}_t\}$.

2) Set ρ at some fixed value ρ_{set}, for example, 0.9999.

3) Calculate $\hat{\phi} = \frac{\hat{\sigma}_{\hat{u}\hat{v}}}{\hat{\sigma}^2_{\hat{v}}} = \frac{\sum_{t=1}^{n} \hat{u}_t \hat{v}_t}{\sum_{t=1}^{n} \hat{v}_t^2}$. Then the bias-corrected estimated predictive coefficient is $\hat{\beta}_L = \hat{\beta} - \hat{\phi}(\hat{\rho} - \rho_{set})$.

4) Estimate the variance of $\hat{\beta}_L$ by $\hat{\sigma}^2_{\hat{w}}(X'X)^{-1}_{(2,2)}$, where $\hat{w}$ is the residual from a regression of $\hat{u}_t$ on $\hat{v}_t$ and $\hat{\sigma}^2_{\hat{w}} = (\frac{1}{n-3} \sum_{t=1}^{n} \hat{w}_t^2)$, the first column of X is a vector of ones, and the second column of X is $(x_0, ..., x_{n-1})'$.

5) Using $\hat{\beta}_L$ and the square root of its estimated variance, the t-statistic is calculated and used for hypothesis testing.

This methodology performs well in a one-sided test for the monthly time series analyzed in Lewellen's (2003) paper, which have $\hat{\rho}$ close to 0.9999.

A.3 A Modified Lewellen Method (L^M)

As we see in the simulations later, Lewellen's method performs well when the true value of ρ is near unity, as assumed. The question is whether one can assume a value for ρ_{set} which depends on the data instead of setting it to be 0.9999. A modified Lewellen method is therefore proposed here by first setting $\rho_{set} = \hat{\rho}^c = (1/n + 3/n^2) + (1 + 3/n + 9/n^2)\hat{\rho}$, a bias-corrected estimator of ρ (see Amihud and Hurvich (2004)), and using Lewellen's method thereafter.

A.4 Augmented Regression Method (ARM)

Hypothesis testing based on the augmented regression, proposed by Amihud and Hurvich (2004)[3], is briefly summarized as follows,

1) The augmented regression model can be written as: $y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t$, where $\{e_t\}$ is independent of both $\{v_t\}$ and $\{x_t\}$. Clearly, by comparing with formula (1), we obtain: $u_t = \phi v_t + e_t$, where $\phi = \sigma_{uv}/\sigma_{v^2}$.

2) Perform an OLS regression of x_t on x_{t-1} to obtain $\hat{\theta}$ and $\hat{\rho}$. Then compute the bias-corrected estimate of ρ,

$$\hat{\rho}^c = (1/n + 3/n^2) + (1 + 3/n + 9/n^2)\hat{\rho}.$$

[3] See also Amihud (2002) for use of this method.

3) Calculate proxies $\{v_t^c\}_{t=1}^n$ for the autoregressive errors $\{v_t\}_{t=1}^n$ using $\hat{\theta}$ and $\hat{\rho}^c$ in equation (2), that is, $v_t^c = x_t - \hat{\theta} - \hat{\rho}^c x_{t-1}$.

4) Obtain the bias-corrected estimators $\hat{\beta}^c$ and $\hat{\phi}^c$ as the coefficients in an OLS regression of y_t on x_{t-1} and v_t^c, respectively, together with a constant.

5) The estimated standard error for $\hat{\beta}^c$ is,

$$\widehat{SE}^c(\hat{\beta}^c) = \sqrt{\{\hat{\phi}^c\}^2 \widehat{Var}(\hat{\rho}^c) + \{\widehat{SE}(\hat{\beta}^c)\}^2}$$

where $\widehat{Var}(\hat{\rho}^c) = (1 + 3/n + 9/n^2)^2 \widehat{Var}(\hat{\rho})$, $\widehat{Var}(\hat{\rho})$ is obtained from the OLS regression in Step (2), and $\{\widehat{SE}(\hat{\beta}^c)\}^2$ is the estimated variance of $\hat{\beta}^c$ in the OLS regression in Step (4). Finally, perform an ordinary t-test based on $\hat{\beta}^c$ and $\widehat{SE}^c(\hat{\beta}^c)$.

B Simulations and Comparisons

The performance of these hypothesis testing methods is investigated and compared in a simulation study, using 1500 simulated replications (data sets) from the model (1) and (2). The parameter values used in the simulation study are estimated values obtained from an actual data set, the predictive regression of the quarterly market return ($NYSE$ value-weighted) on lagged earning-price ratio (Section II.C below). We use the values of the estimated parameters $\hat{\beta}^c$, $\hat{\phi}^c$ and $\hat{\rho}^c$ as if they were the true parameter values. The sample size is $n = 154$, and these parameter values are $\beta = 0.1329$, $\rho = 0.9821$ and $\phi = -3.28$. We construct $u_t = \phi v_t + e_t$, where $\{v_t\}$ and $\{e_t\}$ are mutually independent $i.i.d.$ normal random variables whose standard deviations are 0.02046 and 0.04017, respectively. The results are summarized in Table 1. One-sided (that is, right-tailed) and two-sided hypothesis tests are performed at nominal significance levels 1%, 5% and 10%. In all cases, the null hypothesis is $H_0 : \beta = 0.1329$.

INSERT TABLE 1 HERE

B.1 Bootstrapping (BS)

Using the procedure described in II.A.1, we generate 1500 simulated data sets. For each data set, bootstrapping gives the empirical distribution for $\hat{\beta}_A^*$ under H_0, based on 2500 bootstrap replications.

As shown in Panel A of Table 1, for hypothesis testing with nominal sizes of 1%, 5% and 10%, the resulting one-sided tests have actual sizes close to 10%, 15% and 20% for three bootstrapping procedures: nonparametric (BS^N), parametric-fixed (BS^{Pf}) and parametric-random (BS^{Pr}). The resulting sizes for the corresponding two-sided tests are even bigger. These large distortions in test size imply that the null hypothesis is rejected much more often than it should be.

B.2 Lewellen's Method (L)

In Panel A of Table 1, we used $\rho_{set} = 0.9821$, (the true simulation parameter), as well as the values 0.9721, 0.99, 0.999 and 0.9999.

When ρ_{set} is the true parameter value, 0.9821, the resulting observed test sizes naturally equal the nominal sizes, for both the one-sided and two-sided tests[4]. However, this

[4]The differences between the sizes in the simulations and the nominal sizes are due to simulation errors.

test is infeasible since the researcher does not know the true parameter value. When ρ_{set} differs even slightly from the true parameter value, the performance of Lewellen's tests is weaker.

For the one-sided test, the test sizes are conservative, as pointed out by Lewellen, which implies that the null is rejected less than implied by the nominal size of the Type I error. The size distortion in Lewellen's (2003) test are large for the two-sided test. When ρ_{set} is 0.99, 0.999 and 0.9999, the observed test sizes become, respectively, 2.4%, 11.5% and 13.3% for the nominal 1% test; 9.9%, 26.5% and 29.0% for the nominal 5% test and 16.4%, 38.5% and 40.9% for the nominal 10% test. This means that the null is rejected too often even when the true ρ is only slightly smaller than the assumed ρ. Clearly, the performance of Lewellen's hypothesis testing depends on the assumed value for ρ_{set} relative to the true ρ. In general, the size distortion increases monotonically and quite steeply in $|\rho_{set} - \rho|$.[5] But, if $\rho_{set} < 0.9821$ (0.9721 is used here), Lewellen's test inflates the size.

For our suggestion of a modified Lewellen's method (L^M), where $\rho_{set} = \hat{\rho}^c$, the observed test sizes at nominal 1%, 5% and 10% levels are 11.5%, 20.1% and 25.2% for the one-sided test and 23.1%, 37.0% and 45.4% for the two-sided test. This occurs because using the estimated $\hat{\rho}^c$ as the true value ignores the variability of $\hat{\rho}^c$, reducing the standard error and thus rejecting the null too often.

[5] A detailed set of simulation results of this statement is available upon request.

B.3 Augmented Regression Method (ARM)

For the same simulated data sets, we constructed the augmented regression estimate $\hat{\beta}^c$ and its standard error, based on the procedure described in section II.A.4, and then carried out hypothesis testing using a t-test. The results are indicated by t^{ARM}. Both one-sided and two-sided tests produce observed sizes with reasonably small distortions. For the one-sided test, the observed sizes are almost equal to the nominal sizes. The somewhat inferior results for the two-sided tests are presumably due to asymmetry in the distribution of $\hat{\beta}^c$.

B.4 Power Comparison

Table 1, Panel B shows the power for the bootstrapping (BS), the augmented regression method (ARM) and Lewellen's method (L). The power is calculated by simulating model (1) and (2) for several values of β, centered around the true value of 0.1329.

Ideally, for the one-sided test, where the null hypothesis is $\beta = \beta_0 = 0.1329$ and the alternative to the null is $\beta > \beta_0$, the power of the test should be zero if $\beta < \beta_0 = 0.1329$. At the true $\beta = 0.1329$, the power should equal the nominal size, and for $\beta > 0.1329$ the power of the test should rise steeply to 1.0. We see that the actual power of the ARM t-test is much closer to the ideal pattern than the power of the bootstrap test.

For the two-sided test, the ideal pattern for the power function is as follows: at $\beta = \beta_0 = 0.1329$, the power should equal the nominal size, and for $\beta \neq \beta_0 = 0.1329$, it should rise steeply to 1. We see that the ARM test generally outperforms the bootstrap test. It has similar power as the bootstrapping method when $\beta < \beta_0$, and much higher power when $\beta > \beta_0$. As noted above, the observed size at $\beta = \beta_0 = 0.1329$ is far more accurate for the *ARM* test than for the bootstrap test.

The power of Lewellen's test depends on ρ_{set} and it always rises quickly to 1 because ρ is assumed to be known with certainty, which leads to a small standard error of $\hat{\beta}_L$ (the *ARM* accounts for the uncertainty about $\hat{\rho}^c$). For the one-sided test, if ρ_{set} is larger than the true ρ, the test outperforms the *ARM* with smaller size and better power. But if ρ_{set} is smaller than the true ρ, Lewellen's test is severely oversized. For the two-sided Lewellen's test, even a small difference (in either direction) between ρ_{set} and the true parameter value leads to inflation of the size. Of course, if ρ_{set} were equal to the true parameter value, then both the size and power would be excellent, but this scenario is infeasible because the researcher does not have foreknowledge of the true value of ρ.

B.5 Summary of the Three Methods of Hypothesis Testing

The *ARM* outperforms bootstrapping under all nominal sizes studied for both one-sided and two-sided tests. The superiority of the *ARM* holds for both size and power.

In evaluating Lewellen's (2003) method, comparing his t-test to the one based on the *ARM*, we observe the following. The numerator of Lewellen's t-statistic, the estimated $\hat{\beta}_L$ using $\rho_{set} \approx 1$ is smaller than $\hat{\beta}^c$ (or the true β) if the product $\hat{\phi}(\hat{\rho}^c - \rho_{set}) > 0$; the denominator provides an estimated standard error that is smaller than that of $\hat{\beta}^c$ because Lewellen's method assumes that ρ is known with certainty (being equal to ρ_{set})[6].

[6] Proof is available upon request.

Under a right-tailed test, the resulting size turns out to be smaller than the nominal size, which may be satisfactory to a researcher wishing to consider whether $\{x_t\}$ has predictive power. In addition, Lewellen's test has better power than that of the ARM, perhaps due to the smaller assumed standard error of $\hat{\beta}_L$ under Lewellen's method, although in a range immediately straddling the β under the null hypothesis, the ARM has better power. However, if $\hat{\phi}(\hat{\rho}^c - \rho_{set}) < 0$, $\hat{\beta}_L$ is larger than $\hat{\beta}^c$ or the true β and the denominator is still smaller than $\widehat{SE}^c(\hat{\beta}^c)$, leading to an over rejection of the null hypothesis. This can happen when $\rho_{set} = 0.9721$ (see Table 1, Panel A) or when $\rho_{set} = 0.9999$ and $\hat{\phi} > 0$.

In general, Lewellen's test works very well for right-tailed test if $\hat{\phi}(\hat{\rho}^c - \rho_{set}) > 0$. But the ARM outperforms it in a right-tailed test if $\hat{\phi}(\hat{\rho}^c - \rho_{set}) < 0$, and in all left-tailed tests for $\rho > 0$.[7] The ARM also outperforms Lewellen's method in two-sided tests. In addition, the ARM offers a *unified* approach to both estimation and testing: it provides a method for estimating β and its estimated standard error when ρ is unknown. This standard error can then be used in a t-test. Lewellen's (2003) method separates the testing and the estimation processes. It uses ρ_{set} to calculate a conservatively-estimated $\hat{\beta}_L$ and its standard error to test the hypothesis that $\beta > constant$. If the null is rejected, the estimation of β can be done by the use of an adjusted estimate of ρ from $\{x_t\}$.

[7]If the AR process of x_t is such that $\rho < 0$, other conditions apply for right-tailed tests to provide good performance under Lewellen's method.

C Empirical Illustration

We illustrate the three methods of hypothesis testing by the predictive effect of lagged earning-price ratio, EP_{t-1}, on $RMVW_t$, the value-weighted market return of $NYSE$ stocks. The time period is from the third quarter of 1963, when the earning-price ratio data became available, to the fourth quarter of 2001, total of 154 quarters.[8] $RMVW_t$, the quarterly return, is regressed on EP_{t-1}, the earning-price ratio in the last month of the preceding quarter. Results are presented in Table 2.

INSERT TABLE 2 HERE

OLS regressions of equations (1) and (2) give significant estimators at 5% level for both one- and two-sided tests: $\hat{\beta} = 0.2169$ and $\hat{\rho} = 0.9565$. But we know that both $\hat{\beta}$ and $\hat{\rho}$ are biased, according to Stambaugh (1999) and Kendall (1954). Indeed, we obtain $\hat{\beta}_A = 0.1328$ and $\hat{\beta}^c = 0.1329$.

C.1 Bootstrapping (BS)

Using the procedure described in Section II.A.1, we set the null hypothesis at $\beta_0 = 0$ and run 2500 iterations. Overall, 2500 values of $\hat{\beta}^*_A$ under the null hypothesis are generated for each bootstrapping procedure. For all three procedures (nonparametric, parametric-fix and parametric-random), the mean of $\hat{\beta}^*_A$ is negative. The empirical p-value for the observed $\hat{\beta}_A = 0.1328$ is 0.000 for all procedures and for both one-sided and two-sided tests. All three bootstrapping procedures reject the null hypothesis and show a highly

[8]Data on EP is kindly obtained from J. Lewellen.

significant predictive power of lagged EP on $RMVW$.

C.2 Lewellen's Method (L)

The hypothesis testing for β gives t-statistics of 1.674 (p-value=0.048, one-sided) if $\rho_{set} = 0.999$, and 1.610 (p-value=0.055, one-sided) if $\rho_{set} = 0.9999$ (Recall that $\hat{\rho}^c = 0.9821$). Again, we observe that the test results are sensitive to the assumed ρ_{set}. For the one-sided test, lagged EP is a significant predictor of $RMVW$ at 5% level if $\rho_{set} = 0.999$, but it is insignificance if $\rho_{set} = 0.9999$.

C.3 Augmented Regression Method (ARM)

The augmented regression gives $\hat{\rho}^c = 0.9821$ and $\hat{\beta}^c = 0.1329$ with $\widehat{SE}^c(\hat{\beta}^c) = 0.09102$. The t-statistics for $\hat{\beta}^c$ is 1.460, meaning that the null hypothesis is not rejected at 5% level for either the one-sided or two-sided test. That is, lagged EP has no significant predictive power for $RMVW$. This conclusion does not depend on any foreknowledge of the autoregressive coefficient of the predictive variable and it does not require intensive calculation.

III Multi-Predictor Regression

We propose a hypothesis testing procedure in multi-predictor regression. This complements the model estimation procedure for the multi-predictor case proposed in Amihud and Hurvich (2004, Section 4), based on the Multiple Augmented Regression Method ($mARM$).

We assume that the predictor variables constitute a p-dimensional vector time series $\{x_t\}$ which follows a stationary Gaussian vector autoregressive $VAR(1)$ model. The overall model is given for $t = 1, \dots, n$ by

$$\begin{aligned} y_t &= \alpha + \beta' x_{t-1} + u_t, \\ x_t &= \Theta + \Phi x_{t-1} + v_t, \end{aligned}$$

where we define $(p \times 1)$ vectors,

$$x_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{pt} \end{pmatrix}, \Theta = \begin{pmatrix} \Theta_1 \\ \vdots \\ \Theta_p \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, v_t = \begin{pmatrix} v_{1t} \\ \vdots \\ v_{pt} \end{pmatrix},$$

and a $(p \times p)$ matrix[9],

$$\Phi = \begin{pmatrix} \Phi_{11} & \cdots & \Phi_{1p} \\ \vdots & \ddots & \vdots \\ \Phi_{p1} & \cdots & \Phi_{pp} \end{pmatrix},$$

The quantities y_t, α and u_t are scalars. The vectors $(u_t, v_t')'$ are i.i.d. multivariate normal with mean zero. We allow u_t and v_t' to be contemporaneously correlated and

[9] For a diagonal Φ, Amihud and Hurvich (2004) propose a simpler hypothesis testing method which is similar to the single-predictor testing method.

assume that the absolute values of the eigenvalues of Φ are all less than 1 to ensure stationarity of $\{x_t\}$.

As shown in Amihud and Hurvich (2004), there exists a $(p \times 1)$ vector ϕ and a sequence $\{e_t\}_{t=1}^n$ such that

$$u_t = \phi' v_t + e_t$$

and we can write,

$$y_t = \alpha + \beta' x_{t-1} + \phi' v_t + e_t, \tag{3}$$

where $\{e_t\}$ are i.i.d. normal random variables with mean zero, and is independent of both $\{v_t\}$ and $\{x_t\}$, $Var(u_t) = \sigma_u^2$, $Cov(v_t) = \Sigma_v$, $Var(e_t) = \sigma_e^2 = \sigma_u^2 - \phi' \Sigma_v \phi$.

If the true $\{v_t\}$ is used in the above regression (model (3)), the OLS estimate of β is unbiased. Since $\{v_t\}$ is unknown, it is substituted by the proxy $\{v_t^c\}$, the residuals from a fitted $VAR(1)$ model for $\{x_t\}$, using a bias-correction method that produces a reduced-bias estimation of β. Amihud and Hurvich (2004) propose a bias-corrected estimator of the matrix Φ using a bias expression due to Nicholls and Pope (1988). The procedure is briefly summarized as follows:

1) Use the Nicholls and Pope (1988) expression for the bias of the OLS estimator $\hat{\Phi}$,

$$E[\hat{\Phi} - \Phi] = -b/n + O(n^{-3/2}) \quad ,$$

where,

$$b = \Sigma_v \left[(I - \Phi')^{-1} + \Phi'(I - \Phi'^2)^{-1} + \sum_{\lambda \in Spec(\Phi')} \lambda (I - \lambda \Phi')^{-1} \right] \Sigma_x^{-1} \quad ,$$

I is a $p \times p$ identity matrix, $\Sigma_x = Cov(x_t)$, the symbol λ denotes an eigenvalue of Φ' and the notation $\lambda \in Spec(\Phi')$ means that the sum is for all p eigenvalues of Φ' with each term repeated as many times as the multiplicity of λ. We estimate the bias b iteratively by repeatedly plugging in preliminary estimates of Φ and Σ_v.[10]

2) The preliminary estimator of Σ_v is obtained as the sample covariance matrix of the residuals $(x_t - \hat{\Theta} - \hat{\Phi}x_{t-1})$, if $\hat{\Phi}$ has all its eigenvalues smaller than 1 ($\hat{\Theta}$ and $\hat{\Phi}$ are the OLS estimator of the respective matrices). Otherwise, the Yule-Walker estimator, which is guaranteed to satisfy the stationarity condition, is used,

$$\hat{\Phi}^{YW} = \left[\sum_{t=1}^{n}(x_t - \overline{x}^*)(x_{t-1} - \overline{x}^*)'\right]\left[\sum_{t=0}^{n}(x_t - \overline{x}^*)(x_t - \overline{x}^*)'\right]^{-1}$$

where $\overline{x}^* = \frac{1}{n+1}\sum_{t=0}^{n} x_t$.

3) The bias b is estimated iteratively. In the k^{th} iteration, $\hat{\Phi}^{(k-1)}$ and $\hat{\Sigma}_v^{(k-1)}$, which are the results from the previous iteration, are used to construct $\hat{b}^{(k)}$. Then, $\hat{b}^{(k)}$ is used to calculate $\hat{\Phi}^{(k+1)} = \hat{\Phi}^k + \hat{b}^{(k)}/n$ and $\hat{\Sigma}_v^{(k+1)}$ from the residuals of x_t using $\hat{\Phi}^{(k+1)}$. This iteration process terminates if either the current $\hat{\Phi}^{(k)}$ corresponds to a non-stationary model or a preset maximum of K (in our case, K=10) iterations is reached. We obtain from the last iteration $\hat{\Phi}^c = \hat{\Phi} + \hat{b}/n$.

4) The corrected residual series $\{v_t^c\}$ is constructed using,

$$v_t^c = x_t - (\hat{\Theta}^c + \hat{\Phi}^c x_{t-1}) \quad , \quad t = 1, \ldots, n \quad ,$$

[10]Empirically, we find that the difference for the estimated bias b using iterative or non-iterative estimation is small. Therefore, the theoretical covariance matrix of b is well approximated with one iteration.

where $\hat{\Theta}^c = \bar{x}_t - \hat{\Phi}^c \bar{x}_{t-1}$, the bar indicating sample mean.[11]

The bias-corrected parameter estimates are $\hat{\alpha}^c$, $\hat{\beta}^c$, and $\hat{\phi}^c$, the OLS estimators from the augmented regression of y_t on a constant, all $x_{j,t-1}$ and $v^c_{j,t}$ $(j = 1, \ldots, p)$, respectively.

A Estimation of $Cov[\hat{\beta}^c]$

For hypothesis testing in the multiple predictor case, we need a low-bias estimator of the covariance matrix, $Cov[\hat{\beta}^c]$. We cannot use the estimated *OLS* covariance matrix from the augmented regression, as there is no theoretical justification for this, and indeed our simulations indicate that it produces extremely inaccurate results. We thus use the following result, for constructing an estimator of $Cov[\hat{\beta}^c]$.

Lemma 1

$$E[(\hat{\beta}^c - \beta)(\hat{\beta}^c - \beta)'] = E[(\hat{\Phi}^c - \Phi)'(\phi\phi')(\hat{\Phi}^c - \Phi)] + E[B] \tag{4}$$

where $\hat{\beta}^c$ is the reduced-bias estimator of β obtained from the mARM, ϕ is defined in the multi-predictor model (3), $\hat{\Phi}^c$ is any estimator of Φ based on $x_0, \ldots, x_n$, p is the number of predictors and B is a symmetric $(p \times p)$ matrix given by

$$B = \begin{pmatrix} \frac{(\sum_{t=1}^n r_{1t}e_t)^2}{(\sum_{t=1}^n r_{1t}^2)^2} & \cdots & \frac{(\sum_{t=1}^n r_{1t}e_t)(\sum_{s=1}^n r_{ps}e_s)}{(\sum_{t=1}^n r_{1t}^2)(\sum_{s=1}^n r_{ps}^2)} \\ \vdots & \ddots & \vdots \\ \frac{(\sum_{t=1}^n r_{pt}e_t)(\sum_{s=1}^n r_{1s}e_s)}{(\sum_{t=1}^n r_{pt}^2)(\sum_{s=1}^n r_{1s}^2)} & \cdots & \frac{(\sum_{t=1}^n r_{pt}e_t)^2}{(\sum_{t=1}^n r_{pt}^2)^2} \end{pmatrix}.$$

Here, r_{jt} $(j = 1, \ldots, p)$ are the residuals from an OLS regression of the j'th entry of x_{t-1} on all other $(p-1)$ entries of x_{t-1} as well as all p entries of v^c_t and an intercept, with $v^c_t = x_t - \hat{\Phi}^c x_{t-1} - \hat{\Theta}$.

[11]The *OLS* estimator $\hat{\Theta}$ can be used without correction because the bias of $\hat{\beta}^c$ only depends on $\hat{\Phi}$, but then $\hat{\alpha}^c$ is biased.

Proof: See appendix.

Lemma 2

$$E[B_{i,j}] = \sigma_e^2 E\Big[\frac{\sum_{t=1}^n r_{it} r_{jt}}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big], \;\; i, j = 1, \ldots, p \tag{5}$$

$$E[B_{j,j}] = E\Big[\widehat{SE}(\hat{\beta}_j^c)\Big], \;\; j = 1, \ldots, p \tag{6}$$

where B, $\hat{\beta}_j^c$ and r_{jt} are defined in Lemma 1, σ_e^2 is the variance of e_t in Equation (3), $\widehat{SE}(\hat{\beta}_j^c)$ is the estimated standard error of $\hat{\beta}_j^c$ from the augmented OLS regression of y_t on $x_{1,t-1}$, ..., $x_{p,t-1}$ and v_{1t}^c, ..., v_{pt}^c with intercept: $\widehat{SE}(\hat{\beta}_j^c) = \frac{\hat{\sigma}_e^2}{\sum_{t=1}^n r_{jt}^2}$, where $\hat{\sigma}_e^2$ is the OLS estimate of σ_e^2.

Proof: See appendix.

A feasible approximation for $Cov[\hat{\beta}^c]$ is proposed in the next section.

B Implementation

In the implementation, we make the approximation that $E(\hat{b})$ is b, which is reasonable since $\hat{\Phi}^c$ is a low-bias estimator of Φ. Then, we can approximate $Cov[\hat{\beta}^c]$ by $E[(\hat{\beta}^c - \beta)(\hat{\beta}^c - \beta)']$ and approximate $E[(\hat{\Phi}^c - \Phi)'(\phi\phi')(\hat{\Phi}^c - \Phi)]$ by $Cov[(\hat{\Phi}^c)'\phi]$ (see equation (4)). Note that the left-hand side of equation (4) is not $Cov[\hat{\beta}^c] = E\Big[\big(\hat{\beta}^c - E(\hat{\beta}^c)\big)\big(\hat{\beta}^c - E(\hat{\beta}^c)\big)'\Big]$ because $\hat{\beta}^c$ is a biased estimator of β (although the bias of $\hat{\beta}^c$ is much smaller than that of $\hat{\beta}$). We use the following feasible estimate[12] of $Cov[\hat{\beta}^c]$,

$$\widehat{Cov}^c[\hat{\beta}^c] = \widehat{Cov}[(\hat{\Phi})'\hat{\phi}] + \widehat{E[B]}$$

where $\widehat{Cov}[(\hat{\Phi})'\hat{\phi}]$ and $\widehat{E[B]}$ are estimates of $Cov[(\hat{\Phi})'\hat{\phi}]$ and $E[B]$, corresponding to the formulas,

$$\widehat{Var}^c[\hat{\beta}_j^c] = \sum_{i=1}^{p}(\hat{\phi}_i^c)^2\widehat{Var}[\hat{\Phi}_{ij}] + \sum_{i=1}^{p}\sum_{k\neq i}^{p} 2\hat{\phi}_i^c\hat{\phi}_k^c\widehat{Cov}[\hat{\Phi}_{ij}, \hat{\Phi}_{kj}] + \left[\widehat{SE}(\hat{\beta}_j^c)\right]^2, (j = 1, \ldots, p) \quad (7)$$

$$\widehat{Cov}^c[\hat{\beta}_i^c, \hat{\beta}_j^c] = \sum_{k=1}^{p}\sum_{l=1}^{p}\hat{\phi}_k^c\hat{\phi}_l^c\widehat{Cov}(\hat{\Phi}_{ki}, \hat{\Phi}_{lj}) + \frac{\hat{\sigma}_e^2(\sum_{t=1}^{n} r_{it}r_{jt})}{(\sum_{t=1}^{n} r_{it}^2)(\sum_{t=1}^{n} r_{jt}^2)}, (i, j = 1, \ldots, p) \quad (8)$$

(Equation (7) is a special case of (8) for $i = j$). Here, $\{\hat{\phi}_k^c\}_{k=1}^p$ are the estimated coefficients of $v_{j,t}^c(j = 1, \ldots, p)$ in the multi-predictor augmented regression of y_t on all $x_{j,t-1}(j = 1, \ldots, p)$ and $v_{j,t}^c(j = 1, \ldots, p)$, $v_t^c = x_t - \hat{\Theta}^c - \hat{\Phi}^c x_{t-1}$, which are shown by Amihud and Hurvich (2004, Lemma 4) to be unbiased, $\Big($$\hat{\Phi}^c$ is the reduced-bias $VAR(1)$ coefficient matrix estimated with iterations based on the bias expression due to Nicholls-Pope (1988), as proposed in Amihud and Hurvich (2004)$\Big)$, $\hat{\sigma}_e^2$ is the estimated variance of the error e_t in this regression: $\hat{\sigma}_e^2 = \frac{RSS}{n-(2p+1)}$, where RSS is the residual sum of squares of the augmented regression, $\hat{\Phi}$ is the estimated coefficient matrix of the $VAR(1)$ regression of x_t, obtained by SUR estimation, $\widehat{Var}(\hat{\Phi}_{ij})$ and $\widehat{Cov}(\hat{\Phi}_{ki}, \hat{\Phi}_{lj})$ are the estimated variance and covariance of the coefficients from SUR estimation, $\{r_{jt}\}_{t=1}^n$ for $(j = 1, \ldots, p)$ are the residuals as defined in Lemma 1.

[12]It is possible to estimate $cov[(\hat{\Phi}^c)'\hat{\phi}]$ instead of $cov[(\hat{\Phi})'\hat{\phi}]$, where $\hat{\Phi}^c$ is derived from the Nicholls and Pope (1988) correction formula using a Taylor-expansion (delta method) and the estimated variance-covariance matrix of $\hat{\Phi}$. But using $cov[(\hat{\Phi})'\hat{\phi}]$ to calculate $cov[(\hat{\Phi}^c)'\hat{\phi}]$ greatly simplifies the calculation and it is seen in the simulations that the results are still reliable. Therefore we treat $\hat{\phi}$ as if it were constant and use the estimated variance-covariance matrix of $\hat{\Phi}$ from the seemingly unrelated regression (SUR) method.

It is worth noting that for testing the significance of each $\beta_j, j = 1, \ldots, p$, formula (7) is sufficient. It uses estimates that are directly obtained from most standard statistical software. Formula (8) is necessary for joint tests of all β_j's.

The following is a summary of the procedure for an example of a two-predictor model.

1) Do SUR of x_t on x_{t-1} to obtain $\hat{\Phi}$ and $Cov(\hat{\Phi})$ which is used in (7) and (8). Use $\hat{\Phi}$ to construct a reduced-bias estimator $\hat{\Phi}^c$ using Nicholls and Pope's (1988) bias expression with iterations (see Amihud and Hurvich (2004, Section 4.2)).

2) Construct the bivariate corrected residual series $v_t^c = y_t - \hat{\Theta} - \hat{\Phi}^c x_{t-1}$. Denote $v_t^c = (v_{1,t}^c, v_{2,t}^c)'$ and $x_t = (x_{1,t}, x_{2,t})'$.

3) Obtain $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$ as the coefficients of $x_{1,t-1}$ and $x_{2,t-1}$ in a regression of y_t on $x_{1,t-1}$, $x_{2,t-1}$, $v_{1,t}^c$ and $v_{2,t}^c$, with intercept. This regression also produces $\hat{\phi}_1^c$ and $\hat{\phi}_2^c$ as the coefficients of $v_{1,t}^c$ and $v_{2,t}^c$, and it produces the (2×2) covariance matrix $\widehat{Cov}[\hat{\beta}^c]$ whose diagonal elements are used in (7).

4) Apply formula (7) to get $\widehat{Var}^c[\hat{\beta}_1^c]$, $\widehat{Var}^c[\hat{\beta}_2^c]$ whose squared root values are used for the hypothesis testing of β_1 and β_2.

5) For the joint test of both β_1 and β_2, calculate $\{r_{1t}\}$, the residual from a regression of x_{1t} on x_{2t} and $v_{1,t}^c$, $v_{2,t}^c$, and $\{r_{2t}\}$ accordingly, then apply formula (8).

C Hypothesis Testing

Having both $\hat{\beta}^c$ and an estimate of its covariance matrix, we can proceed to hypothesis testing.

In the $mARM$, individual tests for each predictive variable with $H_0 : \beta_j = 0$ against $H_a : \beta_j > 0$ (one-sided test), or $H_a : \beta_j \neq 0$ (two-sided test) are performed by using the standard t-statistics, calculated from $\hat{\beta}^c_j$ and $\widehat{Var}^c[\hat{\beta}^c_j]$. For the joint test of H_0 : all $\beta_j = 0, j = (1, ..., p)$ against H_a : at least one $\beta_j \neq 0, j = (1, ..., p)$, the standard Wald test is used, employing the matrix $\widehat{Cov}^c[\hat{\beta}^c]$.

D Simulations

D.1 Estimation of $Cov[\hat{\beta}^c]$

To access the performance of our suggested method, we do simulations for a two-predictor case with non-diagonal Φ matrix. $\hat{\beta}^c$ is estimated as in Amihud and Hurvich (2004, Section 5.2.2). In these simulations, we estimate the covariance matrix of $\hat{\beta}^c$, by (7) and (8), and use it in hypothesis testing.

INSERT TABLE 3 HERE

Table 3 presents results for two cases of non-diagonal AR(1) parameter matrix Φ. Case 1:

$$\Phi_1 = \begin{pmatrix} .80 & .1 \\ .1 & .85 \end{pmatrix} \quad ,$$

Case 2:

$$\Phi_2 = \begin{pmatrix} .80 & .1 \\ .1 & .94 \end{pmatrix} \quad ,$$

all with

$$\Sigma_v = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad .$$

For all processes, we generated 1500 simulated replications.

In setting the parameter values of Φ we note that in general, the closer the largest absolute eigenvalue of Φ is to 1, the more nearly non-stationary is the multiple $VAR(1)$ model. Here, the largest absolute eigenvalues of cases 1 and 2 are 0.928 and 0.992, respectively. Case 2 is purposely chosen to study a case that is close to non-stationarity.

We observe first that the OLS estimated standard errors from the augmented regression $\widehat{SE}(\hat{\beta}_1^c)$ and $\widehat{SE}(\hat{\beta}_2^c)$, are much smaller than the true standard deviations of $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$. As expected, $\hat{\phi}_1^c$ and $\hat{\phi}_2^c$ are unbiased (true parameter values are $\phi = (80, 80)$) in all cases here.

Noting that the true parameter values are $(\beta_1, \beta_2) = (0, 0)$. OLS regression gives biased average estimates of 6.47 and 8.24 for $\hat{\beta}_1$ and $\hat{\beta}_2$ for Φ_1 and n = 50. The $mARM$ greatly reduces the average bias to 0.97 and 1.88, respectively. For Φ_2, the OLS bias is greater and also the reduced bias under the $mARM$ is larger, which shows the effect of the near unit root. As expected, when the sample size increases to 200, on average both OLS and the $mARM$ generate much smaller bias, but $mARM$ is still much better: the averages of $\hat{\beta}_1$ and $\hat{\beta}_2$ are 1.24 and 2.13 using OLS, while they are -0.23 and 0.41 using $mARM$—much closer to 0.

Now consider the covariance estimation for $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$. Under Φ_1 with n = 50 and n=200, $\widehat{SE}^c(\hat{\beta}_1^c)$ is 17.086 and 7.804, respectively, which are close to the actual standard deviations of 18.221 and 7.915: the relative errors are only 6.2% and 1.4%. For Φ_2, where the $VAR(1)$ process is close to non-stationary, the relative errors are 5.3% and 1.6%. For β_2, the respective errors under Φ_1 are 8.4% and 0.2% and for Φ_2, they are 14.4% and 6.8%. The estimates $\widehat{Cov}^c[\hat{\beta}_1^c, \hat{\beta}_2^c]$, which are used for the joint test, are also quite close to the actual $Cov[\hat{\beta}_1^c, \hat{\beta}_2^c]$.

Finally, note that $\hat{\phi}_1^c$ and $\hat{\phi}_2^c$ are unbiased, as expected.

D.2 Hypothesis Testing and its Performance

We now examine the test sizes that are obtained by our method, using the model described above. The nominal sizes are 1%, 5% and 10% and the sample sizes are 50 or 200.

We present both individual and joint hypothesis tests based on $mARM$ with the covariance matrix estimated using the method described in section III.A. Both one- and two-sided tests are performed for both the individual-coefficient test, based on t_1 and t_2 for the respective coefficients β_1 and β_2, and we also present the joint Wald test. To evaluate the tests performance, we compare our results with the corresponding tests based on OLS regression of y_t on $x_{1,t-1}$ and $x_{2,t-1}$ with intercept. But in OLS, the coefficients are highly biased. Given the performance of the bootstrap method in the single-predictor case, we do not study it in the *multi*-predictor case.

The results are summarized in Table 4: t_1, t_2 are two individual t-tests with superscript "*one*" and "*two*", referring to the one-sided and two-sided tests. "Wald" is the joint Wald test. We observe the following results from the simulations:

INSERT TABLE 4 HERE

For the individual tests of each coefficient β_1 and β_2, the sizes under $mARM$ tests are uniformly better—closer to nominal—than the simple OLS-based tests. The testing sizes are fairly accurate for our method when the largest absolute eigenvalue of Φ is not too close to 1, and also, naturally, when the sample size is larger.

In the joint hypothesis testing of both β_1 and β_2 being non-zero, we find again that the $mARM$-based test improves on the OLS based test. And again, the sizes are closer to nominal when the largest absolute eigenvalue of matrix Φ is not too close to 1, and of course when n is larger.

E Empirical Illustration

We illustrate multiple hypothesis testing on actual data: predicting $RMVW$, the value-weighted market return of $NYSE$ stocks, by $DIVY$, the dividend yield and by $SPREAD$, yield spread between 3-month treasure bills and BAA bonds (this combines the term spread and the default spread, used in earlier studies). Data are quarterly from 02/1963 to 04/2001, and n=155. Both $DIVY_t$, and $SPREAD_t$ are of the last month of quarter t. This analysis is intended to demonstrate our proposed hypothesis testing methodology for *multi*-predictor regression, rather than to draw any economic conclusions.

We first examine the parameter matrix Φ of the regressors in the vector autoregressive process $VAR(1)$ in a system of equations,

$$\begin{aligned} DIVY_t &= \Phi_{10} + \Phi_{11} DIV_{t-1} + \Phi_{12} SPREAD_{t-1} + v_{1,t} \\ SPREAD_t &= \Phi_{20} + \Phi_{21} DIV_{t-1} + \Phi_{22} SPREAD_{t-1} + v_{2,t} \end{aligned}$$

where $v_{1,t}$ and $v_{2,t}$ are the errors that are serially independent (as obtained from the Durbin-Watson statistics, 1.8066 for $\hat{v}^c_{1,t}$ and 2.0896 for $\hat{v}^c_{2,t}$). We obtain $\hat{\Phi}_{11} = 0.972$, $\hat{\Phi}_{22} = 0.829$, $\hat{\Phi}_{12} = -0.039$ and $\hat{\Phi}_{21} = 0.103$. The Φ_{12} entry in the covariance matrix Φ is significantly non-zero (Φ_{21} is insignificant), which is further confirmed by using Nicholls and Pope's (1988) bias correction as well as by the delta method mentioned in footnote (8).[13] Hence we need to estimate the bivariate augmented regression using the method for non-diagonal Φ matrix.[14] The eigenvalues of this matrix after using the Nicholls-Pope (1988) method are 0.969 and 0.870.

The augmented regression is,

$$RMVW_t = \beta_0 + \beta_1 DIVY_{t-1} + \beta_2 SPREAD_{t-1} + \phi_1 v^c_{DIVY,t} + \phi_2 v^c_{SPREAD,t},$$

where $\{v^c_{DIVY,t}\}$ and $\{v^c_{SPREAD,t}\}$ are the residuals obtained from the bivariate estimation procedure after using Nicholls and Pope (1988) correction for matrix Φ. The estimation results of this augmented regression are presented in Table 5, Panel A. The OLS regression gives $\hat{\beta}_1 = 1.2041$ and $\hat{\beta}_2 = 0.8092$. Both coefficients are significant at 5% level using the t-statistics.

INSERT TABLE 5 HERE

[13] Detailed results from the delta method are available upon request.

[14] If Φ is diagonal, Amihud and Hurvich (2004) presents a simple method of estimating the augmented regression.

However, the OLS predictive regression coefficients $\hat{\beta}$ and the corresponding test statistics are biased in finite samples since we find that the unbiased estimated coefficients ϕ are nonzero: $\hat{\phi}_1^c = -23.854$ and $\hat{\phi}_2^c = -0.507$.

Using the multi-predictor $mARM$, we obtain the reduced-bias coefficients $\hat{\beta}_1^c = 0.8273$, about 70% of $\hat{\beta}_1$ and $\hat{\beta}_2^c = 0.7229$, about 90% of $\hat{\beta}_2$. With the reduced-bias standard errors, both coefficients turn out to be not significant at the 5% level two-sided test. $\hat{\beta}_1$ is not even significant at the 5% level one-sided test. By these results, neither $DIVY$ nor $SPREAD$ have significant predictive effect on the value-weighted market return on the $NYSE$ stocks.

In the joint test, OLS again rejects the null, suggesting that at least one of the predictors has the predictive power, while $mARM$-based test dose not detect significant evidence against the null hypothesis of no effect. The conclusion is consistent with that from the individual coefficient tests.

IV Concluding Remarks

In this paper, we examine predictive regression models where one variable is predicted by other variables that are first-order autocorrelated. We make two contributions:

1) We compare two hypothesis testing methods in single-predictor regression — bootstrapping and Lewellen's method (2003) — with testing based on the augmented regression method (ARM). The results show that ARM-based testing outperforms the other two in most cases, producing more accurate test size and good power. Our method is

outperformed by Lewellen's, which assumes some big ρ_{set}, only in right-tailed tests with $\hat{\phi}(\hat{\rho}^c - \rho_{set}) > 0$. The advantages of the ARM-based testing over other two methods are demonstrated by simulations and an empirical illustration.

2) We propose a convenient new method to estimate the covariance matrix of the estimated predictive coefficients $\hat{\beta}^c$ in multi-predictor regressions and evaluate its performance, using both simulations and an empirical illustration. The individual t-tests, especially the one-sided test, perform well in terms of controlling testing sizes and are more accurate than the simple OLS-based test. Our method also enables *joint* testing of all predictive coefficients using the Wald test which also produces more accurate size than the benchmark OLS based method. The advantage is greater when the sample size is small and/or the coefficient matrix Φ of the regressors vector autoregressive model has eigenvalues close to unit root.

V Appendix

A Proof of Lemma 1

We can write

$$y_t = \tilde{\alpha} + \{\beta' + \phi'(\hat{\Phi}^c - \Phi)\}x_{t-1} + \phi' v_t^c + e_t \quad , \tag{9}$$

where $\tilde{\alpha} = \alpha + \phi'(\hat{\Theta}^c - \Theta)$ is a constant with respect to t. Next, define the $p \times 1$ vectors $\{r_t\}_{t=1}^n$ by $r_t = (r_{1t}, \ldots, r_{pt})'$ where for $j = 1, \ldots n$, $\{r_{jt}\}_{t=1}^n$ is the (row) vector of residuals from a $2p-1$-variable OLS regression of the j'th entry of x_{t-1} on all other entries of x_{t-1} as well as all p entries of v_t^c and an intercept. Correspondingly, define $\{\tilde{r}_t\}_{t=1}^n$ by $\tilde{r}_t = (r_{1t}/\Sigma\, r_{1t}^2, \ldots, r_{pt}/\Sigma\, r_{pt}^2\ )'$ and write $x_t = (x_{1t}, \ldots, x_{pt})'$, and $v_t^c = (v_{1t}^c, \ldots, v_{pt}^c)'$. It follows that

$$\hat{\beta}^c = \sum_{t=1}^{n} \tilde{r}_t y_t \quad , \tag{10}$$

and for all $j, k \in \{1, \ldots, p\}$ with $j \neq k$,

$$\sum_{t=1}^{n} \tilde{r}_{jt} = \sum_{t=1}^{n} \tilde{r}_{jt}\, x_{k,t-1} = \sum_{t=1}^{n} \tilde{r}_{jt}\, v_{jt}^c = \sum_{t=1}^{n} \tilde{r}_{jt}\, v_{kt}^c = 0 \quad , \tag{11}$$

and

$$\sum_{t=1}^{n} \tilde{r}_{jt}\, x_{j,t-1} = \sum_{t=1}^{n} \tilde{r}_{jt}\, r_{jt} = 1 \quad . \tag{12}$$

Substituting y_t from (9) in (10) and using (11) and (12) yields

$$\hat{\beta}^c = \beta + (\hat{\Phi}^c - \Phi)'\phi + \sum_{t=1}^{n} \tilde{r}_t e_t \quad . \tag{13}$$

The Lemma now follows, since e_t has mean 0 and is independent of $\tilde{r}_t$ and $\hat{\Phi}^c$. $\square$

B Proof of Lemma 2

The matrix B is defined as in Lemma 1,

$$B = \begin{pmatrix} \frac{(\sum_{t=1}^n r_{1t}e_t)^2}{(\sum_{t=1}^n r_{1t}^2)^2} & \cdots & \frac{(\sum_{t=1}^n r_{1t}e_t)(\sum_{s=1}^n r_{ps}e_s)}{(\sum_{t=1}^n r_{1t}^2)(\sum_{s=1}^n r_{ps}^2)} \\ \vdots & \ddots & \vdots \\ \frac{(\sum_{t=1}^n r_{pt}e_t)(\sum_{s=1}^n r_{1s}e_s)}{(\sum_{t=1}^n r_{pt}^2)(\sum_{s=1}^n r_{1s}^2)} & \cdots & \frac{(\sum_{t=1}^n r_{pt}e_t)^2}{(\sum_{t=1}^n r_{pt}^2)^2} \end{pmatrix}$$

First, we show the independence between $\{e_t\}$ and $\{r_{jt}\}$.

Because $v_t^c = x_t - \hat{\Theta} - \hat{\Phi}^c$ and $\hat{\Theta}$, $\hat{\Phi}^c$ are all functions of x_t, all entries of v_t^c must be functions of x_t. But e_t is independent of x_t and v_t, so e_t must be independent of v_t^c and therefore independent of r_{it} or r_{jt} for any $i, j = 1, \ldots, p$. Recall that e_t has mean zero. Therefore,

$$\begin{aligned} E[B_{i,j}] &= E\Big[(\frac{\sum_{t=1}^n r_{it}e_t}{\sum_{t=1}^n r_{it}^2})(\frac{\sum_{s=1}^n r_{js}e_s}{\sum_{s=1}^n r_{js}^2})\Big] \\ &= E\Big[\frac{\sum_{t=1}^n \sum_{s=1}^n r_{it}r_{js}e_te_s}{\sum_{t=1}^n r_{it}^2 \sum_{s=1}^n r_{js}^2}\Big] = E\Big[\frac{\sum_{t=1}^n r_{it}r_{jt}e_t^2}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big] \\ &= E\Big[\frac{r_{i1}r_{j1}e_1^2}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big] + \ldots + E\Big[\frac{r_{in}r_{jn}e_n^2}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big] \\ &= E\Big[\frac{r_{i1}r_{j1}}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big]E[e_1^2] + \ldots + E\Big[\frac{r_{in}r_{jn}}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big]E[e_n^2] \\ &= E\Big[\frac{r_{i1}r_{j1}}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big]\sigma_e^2 + \ldots + E\Big[\frac{r_{in}r_{jn}}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big]\sigma_e^2 \\ &= \sigma_e^2 E\Big[\frac{\sum_{t=1}^n r_{it}r_{jt}}{\sum_{t=1}^n r_{it}^2 \sum_{t=1}^n r_{jt}^2}\Big] \end{aligned}$$

which is formula (5).

As a special case, when $i = j$, formula (5) simplifies to,

$$E[B_{j,j}] = E\Big[(\frac{\sum_{t=1}^n r_{jt}e_t}{\sum_{t=1}^n r_{jt}^2})^2\Big] = \sigma_e^2 E\Big[\frac{1}{\sum_{t=1}^n r_{jt}^2}\Big]$$

It remains to be shown that,

$$\sigma_e^2 E\Big[\frac{1}{\sum_{t=1}^n r_{jt}^2}\Big] = E\Big[\frac{\hat{\sigma}_e^2}{\sum_{t=1}^n r_{jt}^2}\Big] \tag{14}$$

Let $H = X(X'X)^{-1}X'$ denote the hat matrix for the regression of y_t on all $x_{j,t-1}$ and v_{jt}^c, $(j = 1, \ldots, p)$, where $X = [1_n, x_{1,t-1}, \ldots, x_{p,t-1}, v_{1t}^c, \ldots, v_{pt}^c]$. Let ϵ denote the residual vector and e the error vector from this regression, so that $\epsilon = (I-H)y = (I-H)e$, where I denotes an $n \times n$ identity matrix. Conditionally on X, we have

$$\sum_{t=1}^n \epsilon_t^2 = e'(I-H)e \sim \sigma_e^2 \chi^2_{n-(2p+1)} \quad ,$$

and since the random variable on the righthand side does not depend on X, the result is true unconditionally as well. Thus,

$$\hat{\sigma}_e^2 = \frac{1}{n-(2p+1)} \sum_{t=1}^n \epsilon_t^2$$

is an unbiased estimator of σ_e^2, that is, $E[\hat{\sigma}_e^2] = \sigma_e^2$. Now, we have

$$\begin{aligned} E\left[\frac{\hat{\sigma}_e^2}{\sum_{t=1}^n r_{jt}^2} \mid X\right] &= E\left[\frac{1}{n-(2p+1)} \frac{e'(I-H)e}{\sum_{t=1}^n r_{jt}^2} \mid X\right] \\ &= \Big(\frac{1}{\sum_{t=1}^n r_{jt}^2}\Big)\Big(\frac{1}{n-(2p+1)}\Big) E[\sigma_e^2 \chi^2_{n-(2p+1)}] \\ &= \sigma_e^2 \frac{1}{\sum_{t=1}^n r_{jt}^2} \quad . \end{aligned}$$

Taking expectations of both sides and using the double expectation theorem yields formula (14). Thus formula (6) is proved. □

References

[1] Amihud, Y. "Illiquidity and Stock Returns: Cross-Sectional and Time-Series Effects." (2002) *Journal of Financial Economics*, 5, 31–56.

[2] Amihud, Y. and C. M. Hurvich. "Predictive Regression: a Reduced-Bias Estimation Method." (2004) *Journal of Financial and Quantitative Analysis*, Forthcoming.

[3] Baker, M. and J. C. Stein. (2003) "Market Liquidity as a Sentiment Indicator." *Journal of Financial Markets*, Forthcoming.

[4] Campbell, J. Y., and Y. Motohiro. (2004) "Efficient Tests of Stock Return Predictability." Working paper, Harvard University.

[5] French, K. R., G.W. Schwert and R.F. Stambaugh. (1987) "Expected Stock Returns and Volatility." *Journal of Financial Economics*, 19, 3–29.

[6] Fuller, W. A. (1996) *Introduction to Statistical Time Series*. New York: Wiley.

[7] Greene, W. H. (2002) *Econometric Analysis*. Prentice Hall.

[8] Hayashi, F. (2000) *Econometrics*. Princeton University Press.

[9] Keim, D. B., and R. F. Stambaugh. (1986) "Predicting Returns in the Stock and Bond Market." *Journal of Financial Economics*, 17, 357–396.

[10] Kendall, M. G. (1954) "Note on Bias in the Estimation of Autocorrelation." *Biometrika* , 41 , 403–404.

[11] Kothari, S. P., and J. Shanken. (1997) "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis." *Journal of Financial Economics* , 169-203.

[12] Lewellen, J. (2003) "Predicting Returns with Financial Ratios." *Journal of Financial Economics*, Forthcoming.

[13] Maynard, A., and K. Shimotsu. (2004) "Covariance-based Orthogonality Tests for Regressors with Unknown Persistence." Working paper, University of Toronto and Queen's University.

[14] Nelson, C. R., and M. J. Kim. (1993) "Predictable Stock Returns: The Role of Small Sample Bias." *Journal of Finance*, 48, 641–661.

[15] Nicholls, D. F., and A. L. Pope. (1988) "Bias in the Estimation of Multiple Autoregressions." *Australian Journal of Statistics*, 30A, 296–309.

[16] Polk, C., S. Thompson and T. Vuolteenaho. (2004) "Cross-sectional Forecasts of the Equity Premium." Working paper, Harvard University.

[17] Stambaugh, R. F. (1986) "Bias in Regressions with Lagged Stochastic Regressors." Working paper, University of Chicago.

[18] Stambaugh, R. F. (1999) "Predictive Regressions." *Journal of Financial Economics*, 54, 375–421.

Table 1: Hypothesis Testing Using Various Methodologies

The tables compare the performances of hypothesis testing under bootstrapping methods (*nonparametric* (BS^N), *parametric-fixed* (BS^{Pf}) and *parametric-random* (BS^{Pr})), the Lewellen's (2003) method (L) under three assumed autoregressive coefficient ($\rho_{set} = 0.9721, 0.9821, 0.99, 0.999$ or 0.9999), a modified Lewellen's method (L^M) and the augmented regression method (ARM), all discussed in the section II.A.

Hypothesis testings performed are both one- and two-sided at sizes 1%, 5% and 10%. Size is the the probability of the type I error.

The parameters are: $\rho = 0.9821$, $\beta = 0.1329$ and $\phi = -3.28$, with $n = 154$, $u_t = \phi v_t + e_t$, where $\{v_t\}$ and $\{e_t\}$ are mutually independent i.i.d. normal random variables whose standard deviation are, respectively, 0.02046 and 0.04017.

The null hypothesis is $H_0 : \beta = 0.1329$. The alternative hypothesis is $H_a : \beta > 0.1329$ for the one-sided test and $H_a : \beta \neq 0.1329$ for the two-sided test.

Panel A: Sizes for the various methods (The differences to the nominal sizes when $\rho_{set} = 0.9821$ are due to the simulation errors).

size	t^{ARM}	BS^N	BS^{Pf}	BS^{Pr}	$L(0.9721)$	$L(0.9821)$	$L(0.99)$	$L(0.999)$	$L(0.9999)$	L^M
					One-sided test					
1%	0.7%	8.9%	11.6%	9.9%	6.7%	0.6%	0.1%	0.0%	0.0%	11.5%
5%	5.0%	13.7%	15.7%	13.0%	20.0%	4.9%	1.4%	0.1%	0.1%	20.1%
10%	10.1%	18.6%	21.0%	17.1%	30.2%	9.1%	2.8%	0.6%	0.5%	25.2%
					Two-sided test					
1%	2.0%	15.5%	18.3%	28.7%	4.5%	0.9%	2.4%	11.5%	13.3%	23.1%
5%	7.9%	26.2%	30.3%	41.7%	12.4%	4.7%	9.9%	26.5%	29.0%	37.0%
10%	13.9%	34.0%	38.5%	50.3%	20.9%	10.5%	16.4%	38.5%	40.9%	45.4%

Panel B: Powers for bootstrapping (BS), the augmented regression method (ARM) and Lewellen's method (L).

In the following table, the first column is the true parameter β used in the data generation processes. The null is $H_0 : \beta = 0.1329$. Subscripts "*1*" and "*2*" indicate one- and two-sided tests, respectively.

1) 1% test,

True β	t_1^{ARM}	BS_1^N	BS_1^{Pf}	BS_1^{Pr}	$L_1(0.9721)$	$L_1(0.9821)$	$L_1(0.9999)$		t_2^{ARM}	BS_2^N	BS_2^{Pf}	BS_2^{Pr}	$L_2(0.9721)$	$L_2(0.9821)$	$L_2(0.9999)$
-0.3987	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		93.7%	80.2%	92.9%	92.7%	100.0%	100.0%	100.0%
-0.2658	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%		81.9%	79.4%	91.2%	91.1%	99.9%	100.0%	100.0%
-0.1329	0.0%	0.1%	0.5%	0.5%	0.0%	0.0%	0.0%		60.5%	71.4%	78.7%	80.2%	96.3%	98.3%	99.7%
0	0.0%	2.1%	3.0%	2.8%	0.0%	0.0%	0.0%		25.5%	49.6%	53.7%	60.1%	42.2%	64.5%	88.7%
0.1329	**0.7%**	**8.9%**	**11.6%**	**9.9%**	**6.7%**	**0.6%**	**0.0%**		**2.0%**	**15.5%**	**18.3%**	**28.7%**	**4.5%**	**0.9%**	**13.3%**
0.2658	12.6%	18.4%	25.7%	20.4%	82.8%	68.0%	29.3%		7.2%	15.9%	22.4%	26.3%	77.7%	61.8%	22.5%
0.3987	83.4%	29.1%	40.7%	31.5%	99.4%	98.1%	92.7%		71.6%	22.5%	35.5%	34.8%	98.7%	97.5%	89.6%
0.5316	99.9%	35.9%	50.3%	39.9%	99.9%	99.9%	99.8%		99.6%	29.6%	44.9%	42.5%	99.9%	99.9%	99.7%
0.6645	100.0%	40.1%	56.2%	44.3%	100.0%	100.0%	99.9%		100.0%	34.3%	49.7%	46.3%	100.0%	100.0%	99.9%

2) 5% test,

True β	t_1^{ARM}	BS_1^N	BS_1^{Pf}	BS_1^{Pr}	$L_1(0.9721)$	$L_1(0.9821)$	$L_1(0.9999)$		t_2^{ARM}	BS_2^N	BS_2^{Pf}	BS_2^{Pr}	$L_2(0.9721)$	$L_2(0.9821)$	$L_2(0.9999)$
-0.3987	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		96.8%	90.9%	97.7%	95.8%	100.0%	100.0%	100.0%
-0.2658	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%		90.3%	91.0%	95.3%	94.6%	100.0%	100.0%	100.0%
-0.1329	0.0%	0.4%	1.1%	1.1%	0.0%	0.0%	0.0%		72.7%	83.3%	84.9%	85.9%	98.7%	99.6%	99.9%
0	0.3%	3.5%	5.3%	4.8%	0.1%	0.0%	0.0%		40.9%	61.5%	64.3%	68.7%	62.2%	80.1%	95.6%
0.1329	**5.0%**	**13.7%**	**15.7%**	**13.0%**	**20.0%**	**4.9%**	**0.1%**		**7.9%**	**26.2%**	**30.3%**	**41.7%**	**12.4%**	**4.7%**	**29.0%**
0.2658	39.2%	33.5%	38.7%	29.3%	93.1%	84.7%	50.5%		25.1%	26.9%	33.5%	36.4%	88.5%	77.3%	40.9%
0.3987	97.5%	46.9%	58.0%	42.3%	99.9%	99.6%	97.5%		94.5%	38.5%	50.3%	46.9%	99.7%	99.2%	96.0%
0.5316	100.0%	54.7%	68.3%	49.5%	100.0%	99.9%	99.9%		100.0%	46.1%	60.7%	53.5%	99.9%	99.9%	99.9%
0.6645	100.0%	61.3%	75.1%	54.7%	100.0%	100.0%	100.0%		100.0%	50.9%	66.1%	56.9%	100.0%	100.0%	100.0%

3) 10% test,

True β	t_1^{ARM}	BS_1^N	BS_1^{Pf}	BS_1^{Pr}	$L_1(0.9721)$	$L_1(0.9821)$	$L_1(0.9999)$		t_2^{ARM}	BS_2^N	BS_2^{Pf}	BS_2^{Pr}	$L_2(0.9721)$	$L_2(0.9821)$	$L_2(0.9999)$
-0.3987	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		97.8%	95.3%	98.9%	96.9%	100.0%	100.0%	100.0%
-0.2658	0.0%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%		93.1%	95.1%	96.1%	95.6%	100.0%	100.0%	100.0%
-0.1329	0.0%	0.9%	1.7%	1.5%	0.0%	0.0%	0.0%		78.3%	87.0%	87.3%	87.7%	99.5%	99.7%	100.0%
0	1.1%	4.9%	6.5%	6.1%	0.1%	0.0%	0.0%		49.5%	66.5%	69.1%	72.7%	71.5%	87.5%	97.5%
0.1329	**10.1%**	**18.6%**	**21.0%**	**17.1%**	**30.2%**	**9.1%**	**0.5%**		**13.9%**	**34.0%**	**38.5%**	**50.3%**	**20.9%**	**10.5%**	**40.9%**
0.2658	57.0%	43.3%	45.9%	35.2%	96.0%	90.7%	63.1%		39.2%	34.1%	40.1%	43.1%	93.1%	84.7%	50.7%
0.3987	99.3%	56.2%	65.1%	46.9%	99.9%	99.9%	98.8%		97.5%	46.9%	58.0%	52.4%	99.9%	99.6%	97.5%
0.5316	100.0%	65.8%	77.7%	54.4%	100.0%	100.0%	99.9%		100.0%	54.7%	68.3%	58.3%	100.0%	99.9%	99.9%
0.6645	100.0%	72.3%	84.9%	60.7%	100.0%	100.0%	100.0%		100.0%	61.3%	75.1%	62.5%	100.0%	100.0%	100.0%

Table 2: Hypothesis Testing: Quarterly Stock Return Predicted by Lagged Earning Price Ratio.
The models are,

$$RMVW_t = \alpha + \beta EP_{t-1} + u_t \quad (1)$$
$$EP_t = \theta + \rho EP_{t-1} + v_t \quad (2)$$

$RMVW$ is the value-weighted $NYSE$ quarterly stocks return, EP is the earning price ratio on the last month of the quarter. The data are from 03/1963 to 04/2001, 154 observations.

Panel A: Ordinary least square (OLS) estimation.

Coefficient	Estimated value	t-statistics
$\hat{\beta}$	0.2169	2.4245
$\hat{\rho}$	0.9565	40.9069

Panel B: Bootstrapping (BS) methods (Section II.A.1). $\hat{\beta}_A = 0.1328$, and we generate 2500 $\hat{\beta}_{boot}$.

Procedure	Mean	Variance	Maximum	Minimum	p-value
nonparametric BS^N	-1.16	0.53	0.26	-6.73	0.000
parametric-fixed BS^{Pf}	-1.47	0.43	-0.38	-5.01	0.000
parametric-random BS^{Pr}	-4.79	10.96	-0.50	-34.91	0.000

Panel C: The Lewellen's method (Section II.A.2)

ρ_{set}	Estimated $\hat{\beta}_L$	t-statistic	p-value (one-sided)
0.99	0.1069	2.3123	0.011
0.999	0.0774	1.6737	0.048
0.9999	0.0744	1.6098	0.055

Panel D: Under the augmented regression method:

$$RMVW_t = \alpha + \beta EP_{t-1} + \phi v_t^c + e_t$$

$\hat{\beta}^c = 0.1329$ with $t = 1.460$ (p-value=0.0732, one-sided), $\hat{\rho}^c = 0.9821$ and $\hat{\phi}^c = -3.28$. v_t^c is obtained using model (2) and $\hat{\rho}^c$.

Table 3: Simulation Results: Parameter Estimates for a Two-predictor Model

1500 replications from the two-predictor models.

$$y_t = \alpha + \beta' x_{t-1} + u_t \quad , \tag{3}$$

$$x_t = \theta + \Phi x_{t-1} + v_t \quad . \tag{4}$$

where β, x_t, θ and v_t are (2×1) matrices. Φ is a (2×2) matrix, i.e.

$$y_t = \alpha + \beta_1 x_{1,(t-1)} + \beta_2 x_{2,(t-1)} + u_t \quad ,$$

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \begin{pmatrix} x_{1,(t-1)} \\ x_{2,(t-1)} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

Table 3 presents estimation results of the two-predictor model by OLS and SUR as well as by the multi-predictor augmented regression method ($mARM$). The five-step estimation procedure is described in Section III.B.

The parameters are: $\alpha = 0$, $\beta = (0,0)'$, $\Theta = (0,0)'$, $u_t = \phi' v_t + e_t$, e_t is $N(0,1)$, $\phi = (\phi_1, \phi_2)' = (-80, -80)'$, v_t is $N(0, \Sigma_v)$. $\{e_t\}$ and $\{v_t\}$ are mutually and serially independent. $n = 50$ or 200.

Two cases are considered for non-diagonal $VAR(1)$ parameter matrices,

$$\Phi_1 = \begin{pmatrix} .80 & .1 \\ .1 & .85 \end{pmatrix} \quad ,$$

and

$$\Phi_2 = \begin{pmatrix} .80 & .1 \\ .1 & .94 \end{pmatrix} \quad ,$$

all with

$$\Sigma_v = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad .$$

Φ_1 has eigenvalues of 0.722 and 0.928, while Φ_2 has eigenvalues of 0.748 and 0.992 (which is close to the unit root).

In the table, $\hat{\beta}_j$ and $\hat{\beta}_j^c$ are the OLS and $mARM$-estimated coefficients, $\widehat{SE}(\hat{\beta}_j)$ and $\widehat{SE}(\hat{\beta}_j^c)$ are the corresponding estimated standard errors, $\hat{\Phi}_{ij}$ are the SUR-estimated $VAR(1)$ coefficients of $\{x_t\}$, $\hat{\phi}_j^c$ is the estimated coefficient of $v_{t,j}^c$ from $mARM$ ($v_{t,j}^c$ is the bias-corrected residual of x_t in the $VAR(1)$ model), $\widehat{SE}^c(\hat{\beta}_j^c)$ and $\widehat{Cov}^c(\hat{\beta}_i^c, \hat{\beta}_j^c)$ are the corrected variance and covariance estimates of $\hat{\beta}^c$ using (7) and (8), the actual standard deviations of the corresponding estimates are denoted by $Std.Dev.$, $TrueCov$ is the true covariance based on 1500 simulations.

Table 3: Results for the two-predictor model (3) and (4)

	(n=50)		(n=200)	
	Φ_1	Φ_2	Φ_1	Φ_2
$\hat{\beta}_1$	6.46716	6.93237	1.24331	1.23748
$\widehat{SE}(\hat{\beta}_1)$	17.62317	17.51045	7.86316	7.83703
$Std.Dev.(\hat{\beta}_1)$	19.87061	19.62003	8.19269	8.25446
$\hat{\beta}_1^c$	0.96890	2.41613	-0.23136	0.01302
$\widehat{SE}^c(\hat{\beta}_1^c)$	17.08633	16.97704	7.80397	7.77803
$Std.Dev.(\hat{\beta}_1^c)$	18.22078	17.92875	7.91485	7.90856
$\widehat{SE}(\hat{\beta}_1^c)$	0.0925	0.0930	0.0402	0.0402
$\hat{\beta}_2$	8.24193	10.17698	2.13111	2.61111
$\widehat{SE}(\hat{\beta}_2)$	16.12600	12.82955	6.95677	4.74827
$Std.Dev.(\hat{\beta}_2)$	18.36973	15.37440	7.13647	5.19485
$\hat{\beta}_2^c$	1.87715	3.03556	0.414517	0.71549
$\widehat{SE}^c(\hat{\beta}_2^c)$	15.63476	12.43873	6.90440	4.71252
$Std.Dev.(\hat{\beta}_2^c)$	17.06413	14.53531	6.91680	5.05692
$\widehat{SE}(\hat{\beta}_2^c)$	0.0846	0.0684	0.0356	0.0244
$\hat{\phi}_1^c$	-79.99782	-79.99842	-79.99776	-79.99771
$\hat{\phi}_2^c$	-80.00081	-80.00023	-80.0025	-80.00249
$\hat{\Phi}_{11}$	0.7110	0.7052	0.7799	0.7794
$\hat{\Phi}_{21}$	0.1082	0.1081	0.1045	0.1052
$\hat{\Phi}_{12}$	0.0939	0.0784	0.0981	0.0931
$\hat{\Phi}_{22}$	0.7530	0.8344	0.8253	0.9142
$\widehat{Cov}^c(\hat{\beta}_1^c, \hat{\beta}_2^c)$	-190.36825	-149.0707	-42.38445	-31.63200
$TrueCov(\hat{\beta}_1^c, \hat{\beta}_2^c)$	-191.11986	-144.61550	-39.95334	-31.19948

Table 4: Simulation Results: Test Sizes for Two-predictor Model

The parameters are the same as in Table 3. 1500 replications from the double-predictor models.

Test sizes are 1%, 5% and 10%. The sizes obtained from the multi-predictor augmented regression method ($mARM$) is compared with the benchmark method: OLS. Two sample sizes of n=50 and 200 are considered. Φ_1 and Φ_2 are the same as described in Table 3.

t_1 and t_2 are the t-statistics for coefficients β_1 and β_2. Wald test is done based on the estimated variance covariance matrix from either OLS or the augmented regression.
The superscript "one" and "two" are used for the one- and two-sided tests.

Table 4: Results for the two-predictor model hypothesis testing

1) 1% test,

Test type			Individual tests				Joint test
			t_1^{one}	t_2^{one}	t_1^{two}	t_2^{two}	Wald
OLS	n=50	Φ_1	4.0%	4.2%	2.7%	2.9%	4.1%
		Φ_2	4.0%	8.4%	2.8%	5.5%	7.5%
	n=200	Φ_1	1.7%	2.4%	1.3%	1.3%	1.7%
		Φ_2	1.9%	4.4%	1.5%	3.1%	4.7%
mARM	n=50	Φ_1	1.9%	1.6%	2.0%	2.0%	3.8%
		Φ_2	1.7%	3.5%	1.9%	2.6%	4.2%
	n=200	Φ_1	1.1%	1.1%	1.1%	0.7%	1.5%
		Φ_2	1.3%	1.9%	0.9%	1.7%	3.8%

2) 5% test,

Test type			Individual tests				Joint test
			t_1^{one}	t_2^{one}	t_1^{two}	t_2^{two}	Wald
OLS	n=50	Φ_1	11.9%	15.3%	9.2%	10.6%	12.3%
		Φ_2	12.9%	23.3%	9.2%	16.1%	21.8%
	n=200	Φ_1	7.3%	8.9%	6.3%	6.1%	7.9%
		Φ_2	7.4%	14.6%	6.7%	9.7%	15.2%
mARM	n=50	Φ_1	6.6%	7.1%	6.9%	6.6%	9.9%
		Φ_2	6.9%	9.9%	6.1%	8.7%	10.8%
	n=200	Φ_1	4.8%	5.5%	6.1%	5.1%	7.0%
		Φ_2	5.1%	7.7%	6.0%	6.4%	10.9%

3) 10% test,

Test type			Individual tests				Joint test
			t_1^{one}	t_2^{one}	t_1^{two}	t_2^{two}	Wald
OLS	n=50	Φ_1	20.1%	25.1%	16.3%	18.8%	21.5%
		Φ_2	20.1%	33.4%	16.8%	25.1%	33.6%
	n=200	Φ_1	12.7%	16.3%	11.5%	12.0%	13.1%
		Φ_2	12.6%	24.3%	12.3%	16.7%	24.3%
$mARM$	n=50	Φ_1	11.0%	13.0%	12.1%	13.0%	16.3%
		Φ_2	12.6%	15.9%	12.1%	14.6%	17.7%
	n=200	Φ_1	9.1%	10.4%	10.5%	9.8%	13.1%
		Φ_2	9.3%	13.6%	11.0%	12.0%	17.5%

Table 5: Hypothesis Testing of a Bivariate Model

The table presents results of the following models:

$$RMVW_t = \alpha + \beta_1 DIVY_{t-1} + \beta_2 SPREAD_{t-1} + u_t \tag{5}$$

$$\begin{pmatrix} DIVY_t \\ SPREAD_t \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix} \begin{pmatrix} DIVY_{t-1} \\ SPREAD_{t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{6}$$

where $RMVW$ is the value-weighted $NYSE$ stocks return, $DIVY$ is the dividend yield and $SPREAD$ is the yield spread between annual yields of 3-month treasure bills and BAA bonds. All data are in the percentages. Returns are quarterly, and the predictive variables are observed on the last month of the lagged quarter. The period is 2/1963 - 4/2001.

The estimated predictive models are:

a) OLS predictive regression: following model (5).

b) Multi-predictor augmented regression method ($mARM$):

$$RMVW_t = \alpha + \beta_1 DIVY_{1,t-1)} + \beta_2 SPREAD_{t-1} + \phi_1 \hat{v}^c_{DIVY,t} + \phi_2 \hat{v}^c_{SPREAD,t} + u_t. \tag{7}$$

The $\hat{v}^c_{DIVY,t}$ and $v^c_{SPREAD,t}$ are the residuals from the $VAR(1)$ regression of $DIVY_t$ and $SPREAD_t$. They are calculated following the procedure described in the reduced-bias method of Nicholls and Pope (1988). $\widehat{Var}^c(\hat{\beta}^c_1)$, $\widehat{Var}^c(\hat{\beta}^c_2)$ and $\widehat{Cov}^c(\hat{\beta}^c_1, \hat{\beta}^c_2)$ are calculated using the formula (7) and (8) in the paper, for both $DIVY$ and $SPREAD$.

Estimate $\hat{\beta}$ is obtained from OLS regressions of model described by (5). Estimate $\hat{\beta}^c$ are obtained from the regression of model described by (7).

In parentheses there are the standard errors of the estimated coefficients and $[t]$ is the corresponding t-statistic.

For individual tests of β_1 and β_2:

Model	Coefficient	$DIVY_{t-1}$	$SPREAD_{t-1}$
OLS	$\hat{\beta}$	1.2041	0.8092
	$(\widehat{SE}[\hat{\beta}])$ $[t]$	(0.5882) [2.05]	(0.3845) [2.10]
$mARM$	$\hat{\beta}^c$	0.8293	0.7231
	$(\widehat{SE}[\hat{\beta}^c])$ $[t]$	(0.2069) [4.01]	(0.1351) [5.35]
	$(\widehat{SE}^c[\hat{\beta}^c])$ $[t]$	(0.5838) [1.42]	(0.3816) [1.89]
	$\hat{\phi}^c$	-23.854	-0.507

For joint test of (β_1, β_2):

Model	Estimated covariance $\widehat{Cov}(\hat{\beta}_1^c, \hat{\beta}_2^c)$	Wald-statistics	p-value
OLS	-0.02262	9.58	0.008
$mARM$	0.00848	5.41	0.067